Barrier for data sharing

Gain is usually with the data consumer, burden is with the data provider

Data owner

Data consumer

Control
Willingness to share data

**ROI**

Return on Investment (ROI) is determined by the balance between effort it takes to share data, and the gain received by sharing data.

**Trust**

Trust is determined by the balance between the risks (due to privacy or competition), and the control (due to verification and security) of sharing and usage of data.

---

Gains Effort

Return on investment

Control Risk

Trust

Original diagram by Nadia Piet, Ocean Conijn, and Joris van Rossum
Working prototype on trusted data sharing

- No access to data itself by Data Consumer
- Control by Data Provider:
  - on request: affiliation, purpose
  - on input: algorithm (code, dependencies, certification)
  - on output: inspection before release
  - on execution: transaction logging, revoke permission, no network access, check on algorithm modifications

Data Provider

Data Consumer (Algorithm Provider)

Data

Trusted Third Party

Secure container

Result

Curation of result

Workflow

Share data

Request

Verify algorithm

Run

Verify output

Release output

Data provider shares data with trusted third party
Data consumer shares algorithm with trusted third party
Data consumer makes request to data provider;
Data provider verifies requester and algorithm;
trusted third party creates secure container;
Data provider verifies output and algorithm behaviour;
Once released, data consumer receives the output.
Data is shared with the Data exchange
Algorithm is shared with the Data exchange by researcher
Researcher makes a request to the data provider
Data provider reviews request and selects dataset
Trusted Third Party runs algorithm on dataset
Data provider reviews output
<table>
<thead>
<tr>
<th>Data Owner</th>
<th>Algorithm Name</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:freek.dijkstra@surfsara.nl">freek.dijkstra@surfsara.nl</a></td>
<td>calculate_sum.py</td>
<td>30717</td>
</tr>
<tr>
<td>Permission Type</td>
<td>Algorithm Dependencies</td>
<td></td>
</tr>
<tr>
<td>one time permission</td>
<td>sys</td>
<td></td>
</tr>
<tr>
<td>Permission Information</td>
<td>Algorithm Length</td>
<td></td>
</tr>
<tr>
<td>The selected algorithm will be ran on the selected dataset of the data owner exactly once.</td>
<td>Lines: 22, Words: 44, Characters: 522</td>
<td></td>
</tr>
<tr>
<td>Choose dataset</td>
<td></td>
<td>random_numbers.txt</td>
</tr>
</tbody>
</table>
Data provider can at any time withdraw permissions
Next Steps

- Better understanding of the needs and requirements
  - This webinar!
- Work with potential pilot partners
  - Integrate with ODISSEI Data Node
  - Talks with other interested organisations (perhaps you?)
- Make the demo accessible to everyone.
  - Currently requires a SURF ResearchDrive account
  - We’ll make it work with e.g. Google Drive (August 2020)
Different Methods to Ease Data Sharing

<table>
<thead>
<tr>
<th><strong>Agreements</strong></th>
<th><strong>Registration</strong></th>
<th><strong>Pseudonymization</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Stipulation of what can/cannot be done</td>
<td>• Authentication</td>
<td>• Filtering (on records)</td>
</tr>
<tr>
<td>• Signing of contract or NDA</td>
<td>• Verification of credential</td>
<td>• Pruning (on properties)</td>
</tr>
<tr>
<td>• Dispute resolution process</td>
<td>• Reputation score</td>
<td>• Aggregation (combine records)</td>
</tr>
<tr>
<td></td>
<td>• Policy framework</td>
<td>• Make coarse grained buckets</td>
</tr>
<tr>
<td></td>
<td>• Audit trails</td>
<td>• Slight alteration of data</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Data Vault</strong></th>
<th><strong>Secure Containers</strong></th>
<th><strong>Secure Computing</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Data source retains control</td>
<td>• Bring algorithm to data</td>
<td>• Secure multi-party computation</td>
</tr>
<tr>
<td>• Delegate permissions</td>
<td>• At Trusted third party or at data provider</td>
<td>• Homomorphic encryption</td>
</tr>
<tr>
<td>• No central data lake</td>
<td>• Share output instead of data</td>
<td>• Garbled Circuits</td>
</tr>
<tr>
<td>• Data marketplace</td>
<td></td>
<td>• Zero-knowledge proof</td>
</tr>
</tbody>
</table>
COLLABORATION WITHOUT SHARING DATA

Freek Dijkstra
Freek.Dijkstra@surfsara.nl
www.surf.nl

This presentation is available under the creative commons attribution 4.0 license

Driving innovation together